Disagreeing about development: An analysis of parent-teacher agreement in ADHD symptom trajectories across the elementary school years

Murray, Aja Louise ; Booth, T ; Ribeaud, Denis ; Eisner, M

Abstract: Objectives: It is well-known that in cross-sectional analyses, agreement between informants is modest as best when rating ADHD and other disruptive behaviour disorder symptoms. We here aimed to develop recommendations for the use of multi-informant data in the context of longitudinal developmental analyses that examine symptom trajectories over time. Method: Using parallel process modelling, we estimated parent-teacher agreement in inattention and hyperactivity/impulsivity symptom initial levels and slopes across the elementary school years (ages 7, 9 and 11) for a community sample of n=1388 youth. We also used these models to examine whether initial levels and slopes differed significantly across informants. Results: Informant agreement was low to moderate and higher for inattention slopes (r=.47) than for hyperactivity/impulsivity slopes (r=.23). Parents and teachers reported opposite developmental trends for inattention with teachers reporting declines and parents reporting increases over time. Parents reported overall higher levels of hyperactivity/impulsivity but there were no average informant differences in slopes. Conclusion: Of the options available, we recommend specifying separate but correlated factors for different informants in developmental analyses of ADHD. This can be achieved within latent growth curve and growth mixture models.

DOI: https://doi.org/10.1002/mpr.1723

Disagreeing about development: An analysis of parent-teacher agreement in ADHD symptom trajectories across the elementary school years

Short title: ADHD cross-informant agreement

Aja Louise Murray[1]

Tom Booth[2]

Denis Ribeaud[3]

Manuel Eisner[1]

[1] Institute of Criminology, University of Cambridge, UK

[2]Department of Psychology, University of Edinburgh, UK

[3]Jacobs Center for Productive Youth Development, University of Zurich, Switzerland

*Corresponding author at: Institute of Criminology, University of Cambridge, CB3 9DA;

email: am2367@cam.ac.uk

Acknowledgements

We are grateful to the children, parents and teachers who provided data for the z-proso study and the research assistants involved in its collection.

Funding sources

Funding from the Jacobs Foundation (Grant 2010-888) and the Swiss National Science Foundation (Grants 100013_116829 & 100014_132124) is also gratefully acknowledged.

Abstract

Objectives: It is well-known that in cross-sectional analyses, agreement between informants is modest as best when rating ADHD and other disruptive behaviour disorder symptoms. We here aimed to develop recommendations for the use of multi-informant data in the context of longitudinal developmental analyses that examine symptom trajectories over time.

Method: Using parallel process modelling, we estimated parent-teacher agreement in inattention and hyperactivity/impulsivity symptom initial levels and slopes across the elementary school years (ages 7, 9 and 11) for a community sample of n=1388 youth. We also used these models to examine whether initial levels and slopes differed significantly across informants.

Results: Informant agreement was low to moderate and higher for inattention slopes (r=.47) than for hyperactivity/impulsivity slopes (r=.23). Parents and teachers reported opposite developmental trends for inattention with teachers reporting declines and parents reporting increases over time. Parents reported overall higher levels of hyperactivity/impulsivity but there were no average informant differences in slopes.

Conclusion: Of the options available, we recommend specifying separate but correlated factors for different informants in developmental analyses of ADHD. This can be achieved within latent growth curve and growth mixture models.

Keywords: Cross-informant agreement, ADHD, latent growth curve modelling

Attention-deficit/hyperactivity disorder (ADHD) is characterised by impairing levels of inattention and/or hyperactivity/impulsivity (APA, 2013). Around 5% of the global population meet diagnostic criteria for ADHD; however, it is generally accepted that ADHD symptoms exist on a continuum, affecting individuals to a greater or lesser extent both above and below clinical cut-offs (e.g. Groen-Blokhuis et al., 2014; Polanczyk et al., 2007). Though traditionally conceived of as a childhood disorder, it is now apparent that there is considerable heterogeneity in the course of ADHD symptoms over development (e.g. Pingault et al., 2011). This insight has been met with a surge in interest in characterising and explaining differences in ADHD symptom developmental trajectories across individuals (e.g. Malone et al., 2010; Moffitt et al., 2015; Murray, Obsuth, Eisner & Ribeaud, 2017a; Van Lier et al., 2007). It is, however, also well known that at a given time point, different informants (e.g. parents vs teachers) provide symptom ratings that are at best moderate in their level of agreement. The extent to which informants also show disagreement on symptom developmental trajectories is, however, yet to be determined. In this study, we thus evaluate parent-teacher agreement in developmental trajectories over ages 7 to 11 in a large longitudinal study of child development. On this basis, we provide recommendations for incorporating multi-informant data into studies of ADHD developmental trajectories.

Developmental analyses concerned with ADHD symptom trajectories and their variation across individuals have been growing in frequency in recent years. In particular, latent growth curve models and growth mixture models have become popular methods of analysing longitudinal ADHD symptom data (e.g. Malone et al., 2010; Döpfner et al., 2015; Jester et al., 2005; Murray et al., 2017a; Van Lier et al., 2007). Using repeated measures of ADHD symptoms over many years, these analyses aim to characterise patterns of ADHD symptom change over time, to establish whether symptom trajectories can be summarised in terms of a small number of ‘developmental subtypes’, and to identify predictors and

outcomes of following particular trajectories. Growth mixture analyses have, for example, provided evidence for a 'late onset' developmental subtype of ADHD characterised by initially low symptom levels in childhood but increases over time (Murray et al., 2017a; Pingault et al., 2011). Amongst other evidence, this has helped to draw attention to the fact that contrary to traditional clinical notions, ADHD is not necessarily an early-onset disorder. Latent growth and growth mixture models have also been used in the context of testing treatment effects (e.g. Swanson et al., 2007; Sonuga-Barke, 2008). Despite their widespread and increasing use, however, there has been little attention paid to the extent to which and in what way developmental analyses are affected by a lack of agreement between informants on ADHD symptoms.

It is known from cross-sectional analyses that cross-informant discrepancies in psychopathology, including ADHD symptoms, abound (e.g. Achenbach, 2005; Kennerley et al., 2016). Discrepancies are also seen in the substantive conclusions of research studies with, for example, significant associations or treatment effects observed according to one informant but not another (e.g. see De Los Reyes, 2013 for an overview). They are also important in clinical contexts where diagnostic criteria require that ADHD symptoms must be evident across multiple contexts for a diagnosis to be made (APA, 2013). However, there are no standardised rules for combining data from informants who observe the child in different contexts (e.g. home, school). The prevalence of cross-informant disagreements thus adds an additional layer of complexity to the already challenging process of diagnosis (e.g. Kennerley et al., 2016).

In the early school years, informant discrepancies most often concern ratings by parents versus teachers. Cross-informant correlations between parents and teachers have ranged from as low as .09 up to around .52 (e.g. see Hartman et al., 2007; Narad et al., 2015 for overviews), representing moderate agreement at best. Various explanations for the lack of

parent-teacher agreement in ADHD symptoms have been proposed including measurement error, systematic biases in the ratings of informants, and differences in behaviour expressed across contexts/in interaction with different informants. Current evidence suggests that all three contribute to low correlations between informants, but perhaps especially the latter (e.g. Hartman et al., 2007; see De Los Reyes, 2013 for a general review).

There have also been some indications that agreement between parents and teachers on ADHD symptoms differs by developmental stage. Using ratings for a self-selected sample (n=6659) of youth aged 4-17, Narad et al. (2015) found that agreement between teachers and parents depended on whether the target child was in the preschool, elementary school, or high school age range. Agreement on inattention was stronger for high school youth (r=.32) as compared to the elementary school group (r=.18), while agreement on hyperactivity/impulsivity was stronger for the preschool group (r=.44) as compared to both the elementary school group (r=. 33) and high school group (r=. 32). However, the study was cross-sectional and the representativeness of the sample unknown, making it impossible to infer generalisable age-related changes in informant discrepancies. Other studies have shown no evidence of age-related differences in informant discrepancies in ADHD symptoms (e.g. Sayal & Goodman, 2009). There are thus hints at but no definitive answers regarding age-related changes in ADHD cross-informant discrepancies.

Low to moderate parent-teacher agreement and potential age differences in informant agreement raise the question of how to best analyse data from multiple informants when conducting developmental analyses in longitudinal data. While there is a consensus that wherever possible, data from multiple informants should be collected, there is less clarity on how to treat multi-informant data in quantitative analyses. There are three broad classes of approaches that can be considered: using data from a single informant, analysing informant data separately, and combining informant ratings. The latter two categories could include a

range of approaches. For example, as regards analysing informant data separately, as well as the basic option of duplicating analyses for each informant, multi-informant data could be analysed separately within the same model. Fitting a parallel process model with separate factors for parent- versus teacher- reports, or identifying classes defined by combinations of parent- and teacher- reports in a growth mixture model would be examples of this approach. As regards combining informant data, this could include analysing a sum or average of scores across the two informants, or it could use a more complex combination method where one informant's rating is weighted more highly for certain symptoms; or where both informants must endorse a symptom for it to contribute to a total symptom count (e.g. Kennerly et al., 2016). Other approaches to combining informant data might include fitting a longitudinal trifactor or multi-trait multi-method model to disentangle informant-unique and informant-general perspectives (e.g. Bauer et al., 2013; Litson et al., 2016).

While each approach has advantages and disadvantages, in the context of developmental analyses, a major factor to consider should be the extent of cross-informant agreement on developmental trajectories. If parents and teachers report markedly different trajectories, it would be difficult to justify approaches that discard data from one informant or blur the distinction between the two. Rather, this situation would call for an approach that can provide the best insight into informant discrepancies in the context of the developmental analysis of interest. While it is clear that parents and teachers show moderate agreement in symptoms cross-sectionally, the extent of agreement on the course of symptoms over time is not known. In this study we, therefore, sought to determine cross-informant agreement in ADHD developmental trajectories as reported by teachers and parents.

Method

Participants

Participants were from the Zurich project on social development from childhood to adulthood (z-proso; http://www.jacobscenter.uzh.ch/en/research/zproso.html). Sample selection involved a stratified random sampling procedure at the level of the school and all children who were to enter the first grade in sampled schools were invited to participate via their parents. Among the N=1675 target sample, n=1388 provided data for the current study which was collected when the children were aged 7, 9 and 11. Comprehensive details of the sample, including recruitment and assessment procedures, and attrition are available at the study website and in previous publications (e.g. Eisner & Ribeaud, 2007; Eisner, Murray, Eisner, Ribeaud, 2018).

Measures

ADHD symptoms were measured using parent- and teacher- versions of the *Social Behavior Questionnaire* (SBQ; Tremblay et al., 1991). This includes four items measuring inattention and four measuring hyperactivity/impulsivity. Z-proso used a German translation of the SBQ but the original English wordings of the items are provided in Table 1. Reponses are provided in a 5-point scale from *never* to *very often*. The questionnaires were administered in paper and pencil form to the teachers, in German, the official language of the study location. In many cases the same teacher completed the questionnaire for a given child at the age 7 and age 9 measurement waves. Between age 9 and 11, classes were reshuffled, thus a different teacher usually completed the questionnaire at age 11. The effects of the teacher change between ages 9 and 11 on ADHD symptom ratings were examined in a previous publication (Murray et al., 2017). The study found that in the context of a growth curve model, the addition of residual covariances between teacher reported data at ages 7, 8, and 9 substantially improved model fit, suggesting that covariation was greater for data provided by the same informant. However, there was no evidence of a qualitative shift in ratings corresponding to the teacher change. This suggests that provided residual covariances

between same-informant ratings over time are included, the teacher change is unlikely to much affect the analysis of developmental trajectories.

Items were administered to parents via interview in their home or another location of their choosing using a computer-assisted personal interview procedure. Interviews were conducted in one of 10 languages, depending on the language of the interviewee. The only other difference was that the teachers completed versions that referred to 'he/she', whereas the interviewers always read out the name of the specific child being referred to. The validity and reliability of the SBQ in the current sample has been supported in several previous studies which have supported the reliability, developmental invariance, and factorial validity of the instrument (e.g. Murray, Eisner & Ribeaud, 2017; Murray, Obsuth, Eisner & Ribeaud, 2017b).

Statistical procedure

Correlations between parent and teacher ratings of the initial levels and change in ADHD over development were estimated using latent growth curve modelling. Inattention and hyperactivity/impulsivity were analysed separately to reflect previous evidence that they differ in developmental trajectories (e.g. Arnold et al., 2014). For each dimension, parallel process models were fit in which latent growth curves were specified for parent-reported data and teacher-reported data. Parent-reported inattention (hyperactivity/impulsivity) and teacher-reported inattention (hyperactivity/impulsivity) were specified as latent factors. In all cases only linear growth was included and parent-reported intercept and linear slope factors were allowed to correlate with the teacher-reported intercept and linear slope factors. Time intervals were specified as proportional to the differences between the median age of the sample at each measurement wave. Scaling and identification were achieved by fixing the mean and variance of the time 1 teacher-reported factor at 0 and 1 respectively and fixing the

loading and intercept of a reference indicator equal across all latent factors. Residual covariances were included between the same items measured over time within informants but not across informants. That is, each teacher-reported item at age 7 was covaried with the corresponding teacher-reported item at ages 9 and 11 and between ages 9 and 11. The same pattern of residual covariances were included for the parent reports.

A χ^2 difference test was used to compare the fit of models in which the intercept factor means were allowed to differ versus constrained to equality across informants. An analogous test was conducted for slope factor means to test their equality across informants. Collectively, this allowed us to test whether parents and teachers differed significantly in their ratings of inattention and hyperactivity/impulsivity initial status and changes over time. All models were estimated in *Mplus 7.13* (Muthén & Muthén, 2014) using maximum likelihood estimation.

Results

Inattention

The parallel process model for inattention showed good fit (CFI=.99, TLI=0.99, RMSEA=.028, SRMR=.026). Key parameters are provided in Table 2 and displayed in Figure 1. Based on these, correlations between teacher and parent intercept factors was *r*=.45. The correlation between teacher and parent slope factors was *r*=.47. Teachers tended to report initially higher levels of inattention than parents and this difference was statistically significant [χ^2 (1) = 21.81, *p*<.001]. In addition, while teachers reported that inattention symptoms decreased on average across ages 7 to 11, parents reported an increase in symptoms on average. This difference in slope was statistically significant [χ^2 (1) = 30.21, *p*<.001].

Hyperactivity/impulsivity

The parallel process model for hyperactivity/impulsivity also showed acceptable fit (CFI=.95, TLI=0.94, RMSEA=.056, SRMR=.046). Key parameters are provided in Table 3 and displayed in Figure 1. The correlation between teacher and parent intercept factors was *r*=.45. The correlation between teacher and parent slope factors was *r*=.23. Teachers tended to report lower initial levels of hyperactivity/impulsivity than parents and this rater difference was statistically significant [χ^2 (1) = 15.67, p<.001]. There was no significant difference in average hyperactivity/impulsivity slopes according to parents versus teachers [χ^2 (1) = 0.339, p=.56]. Both suggested no significant linear increase or decrease in hyperactivity/impulsivity over the studied period.

Discussion

In the current study, we evaluated levels of disagreement on ADHD developmental trajectories across parent and teacher informants. Cross-informant correlations for initial status (intercepts) were *r*=.45 for both inattention and hyperactivity/impulsivity. Cross-informant correlations for linear slopes were *r*=.47 for inattention and *r*=.23 for hyperactivity/impulsivity. On average, while teachers reported an increase in inattention symptoms over ages 7 to 11, parents reported decreases over the same period. There were no significant differences in parent versus teacher reports of hyperactivity/impulsivity changes over development. Teachers reported consistently higher levels than parents but both reported fairly flat symptom trajectories.

Our study adds to the substantial pre-existing literature suggesting that parents and teachers tend to show considerable disagreement on ADHD symptoms in childhood (e.g. Antrop et al., 2002; Murray et al., 2007; Wilcutt et al., 2012; Narad et al., 2015). When compared with cross-sectional estimates of parent-teacher agreement, the initial status correlations of *r*=.45 for both inattention and hyperactivity/impulsivity are towards the upper

end of the inter-rater agreement range reported in previous studies. Given that the SBQ is only a very brief measure of ADHD symptoms, with only 4 items each for inattention and hyperactivity/impulsivity, these levels of agreement suggest that the items are functioning well as a means of measuring a common phenotype across informants. This likely in part reflects the use of a latent variable measurement model, which disattenuates correlations for unreliability; however, another important contributing factor may be the generality of items. None of the SBQ ADHD items refer to any specific context (e.g. the classroom) or specific activities (e.g. schoolwork). There are thus arguably no items where only one of the informants would have sufficient opportunity to observe the behaviour. While 'context-free' items may – on the face of it - be advantageous in promoting stronger cross-informant agreement, the omission of context-specific items risks missing important context-specific behaviours and meaningful differences across contexts (e.g. De Los Reyes, 2011). Moreover, given that exhibiting ADHD across multiple contexts may itself be a marker of severity (e.g. De Los Reyes et al., 2009) we would argue that ADHD measures used in research contexts should measure behaviour across multiple contexts, as well as including context-free items.

The moderate correlations of initial levels occurred in the context of significant mean-level differences between informants on initial status. Previous research has generally found that parents report higher levels of ADHD symptoms (e.g. Antrop et al., 2002; Hart et al.,1995; Kennerley et al., 2016; Murray et al., 2007; Narad et al., 2015; Yeguez & Sibley, 2016), although there have also been some cases where teachers have reported higher levels (e.g. Jester et al., 2005). In the current study, we found that the direction of the difference in initial status ratings depended on the dimension, with teachers reporting higher initial levels of inattention but parents reporting higher initial levels of hyperactivity/impulsivity. Inattention symptoms may be more apparent for teachers who regularly observe children engaged in tasks that tap sustained attention. On the other hand, hyperactive/impulsive

behaviours may be expressed to a greater extent in the home environment, which is generally less structured, may involve more idle time, and which may thus be more evocative of hyperactivity/impulsivity. Such behaviours may also be considered less context-inappropriate in the home, leading to an environment that is more permissive of hyperactivity/impulsivity. The higher levels hyperactivity/impulsivity symptoms observed in the home, however, seem to contradict the sometimes-voiced argument that the school environment may be more challenging for children with high levels of ADHD symptoms and hyperactivity/impulsivity symptoms easier to detect because tasks demands are higher and behavioural expectations stricter (e.g. Hartman et al., 2007; Jester et al., 2005). A previous study in the current sample may help resolve this paradox (Murray, Ribeaud & Eisner, 2018). Using a growth mixture analysis approach, the study found that parent reports of hyperactivity/impulsivity were higher for children with generally low symptom levels but teacher reports were higher for children with elevated symptoms. As the current sample is a normative sample, in the aggregate the pattern in the children with lower levels of symptoms dominates the pattern in the children with higher levels of symptoms.

The opposing results for inattention versus hyperactivity/impulsivity suggest that it would generally be advisable to distinguish between these dimensions when assessing cross-informant differences. Indeed, previous research has indicated that parents and teachers disagree not only on overall levels but also patterns of symptoms. Specifically, parents are more likely to provide ratings that suggest that their child shows problems across both inattention and hyperactivity/impulsivity, whereas teachers are more likely to report significant problems in one domain only (e.g. Kennerley et al., 2016; Sollie et al., 2013; Youngstrom et al., 2000). However, our results also show that mean differences in informant ratings can also depend on developmental stage with systematic differences in reported levels

of inattention largely disappearing by age 11 but for hyperactivity/impulsivity remaining approximately the same.

Our study also provides estimates of cross-informant agreement on the linear change in symptoms over ages 7, 9 and 11. The cross-informant correlation for inattention slope was *r*=.47 but for hyperactivity/impulsivity it was only .23. These are in the range of cross-informant agreements generally observed in cross-sectional data (e.g. Narad et al., 2015). They, however, contradict the idea that agreement in hyperactivity/impulsivity would generally be greater than for inattention because the former is more overt and thus presumably easier to rate (e.g., Kennerley et al., 2016). As there was no evidence that children showed any less variation in hyperactivity/impulsivity slopes than inattention slopes, the explanation must lie elsewhere. It may be due to the fact that children changed teachers and were newly mixed into classrooms between age 9 and 11. It is possible that that hyperactivity/impulsivity symptoms in the school context are more sensitive to changes in the classroom environment. It is known, for example, that classroom factors such as amount of structure, idle time, and stimulation can affect levels of disruptive behaviours in the short term (e.g. Imeraj et al., 2016). A reshuffling of teachers and classes could mean that there was less agreement on hyperactivity/impulsivity slopes than on inattention slopes; on the assumption of the former being more responsive to changes in teachers and classroom environment than the latter. This explanation will require further investigation, through, for example, examination of school and home predictors of change in inattention versus hyperactivity symptoms in the home versus school context. Replication in samples where there were different patterns of teacher change would also help provide a test of this explanation.

Although the cross-informant correlations for inattention intercepts and slopes were in the moderate range, parents and teachers suggested average change in opposite directions for

inattention, the former reporting increases and the latter reporting decreases. The difference in slopes for hyperactivity/impulsivity was not significant. Past research has suggested that there is considerable heterogeneity in ADHD symptom trajectories (e.g. Pingault et al., 2011). However, most studies find that in the aggregate, inattention levels generally remain quite stable while hyperactivity/impulsivity levels tend to decline over the course of development, (e.g. Hart et al., 1995; Lahey et al., 2005). In this respect, the relatively shallow slopes for inattention reported by both teachers and parents (reported levels moving around 0.5 standard deviations over four years) are within the expected range of developmental trajectories. The non-significant hyperactivity/impulsivity slopes are more difficult to map to previous findings. However, they are potentially consistent with concerns that have been raised over the declining applicability of hyperactivity/impulsivity symptom markers over the course of development. Many previous studies have utilised measures based on DSM-IV criteria which refer to symptoms that have an obvious applicability in early childhood (e.g. 'difficulty playing quietly', 'difficulty remaining seated') but which may not adequately capture hyperactivity/impulsivity in late childhood and adolescence. As such, the declines reported in previous studies may partly reflect declining sensitivity to manifestations of hyperactivity/impulsivity with developmental stage (Faraone, Biederman & Mick, 2006). In using a non-DSM based set of items that has been examined for developmental appropriateness at each measurement wave, our measures may have avoided this issue.

It is not clear why parent- and teacher- reports suggested increases versus decreases in inattention respectively. It may reflect the fact that teachers observe and attend primarily to performance on academic tasks. Inattention may decrease over time with increasing mastery of these tasks. In the home environment, however, parents could observe apparent increases in inattention as environmental demands become more complex with age and executive functions are increasingly taxed. To evaluate these speculative interpretations, qualitative

interviews with parents and teachers could be conducted to explore the behaviours and tasks/activities/contexts feeding into informant ratings for children of different developmental stages. Another potential contributing factor could be if a proportion of children with elevated symptoms initiated medication for these symptoms during the time period studied. Medications may be more effective during the school day or not taken at home and during school breaks. This could lead to an improvement in symptoms specific to the school context. However, our self-report data indicate that less than 5% of children were medicated for ADHD, therefore, this could only be part of the picture.

In terms of methodological implications, the modest level of correlation between intercept and slope factors across raters and the average differences across informants imply that parent- and teacher- reported ADHD data should generally be analysed separately. A potentially useful approach is to use a method similar to the parallel process model utilised in the current study. In a growth mixture modelling context, classes could be defined by combinations of parent- and teacher- reports. Alternatively, two latent categorical variables: one for each informant could be modelled and their relation evaluated. These kinds of approaches acknowledge that parents and teachers partly agree and partly disagree and provide information on the extent and nature of the (dis)agreements. They also allow cross-situationality to be captured, which is important, given that showing symptoms across multiple domains is considered a marker of severity over and above symptom counts (e.g. APA, 2013). These models could also be extended to, for example, compare predictors of teacher-reported versus parent-reported slopes. Increases in symptoms in one setting but decreases in another may, for example, be an indicator of a context-specific issue related to, for example, the onset of bullying in the school context or a breakdown of parental relationships in the home context. Finally, future studies could also explore the utility of analysing individual-level discrepancies in linear slopes. One predictor of particular interest

here may be gender. ADHD has traditionally been conceptualised as a male-typical disorder and is associated with less disruptive behaviour when it occurs in females (e.g. Gershon & Gershon, 2002). Teacher-parent agreement could thus be weaker for females in whom symptom manifestations are less well-understood and potentially more subtle. However, females may also show a later onset of symptoms (e.g. Murray, Booth, Eisner, Auyeung, Murray & Ribeaud, 2018), suggesting potential age-by-gender interactions in informant differences and agreement. Consistent with this idea, Van Der Ende & Verhulst (2005) found that the correlation between teacher and parent reports of attention problems was stronger for males in early adolescence but stronger for females in late adolescence in a normative sample of Dutch adolescents. Further research in longitudinal data, particularly that which begins in early childhood, will be required to evaluate whether this pattern generalises.

The level of disagreement between parents and teachers on ADHD symptom trajectories also suggests that developmental studies should not rely on a single informant. The idea of using the data from a single informant is attractive in the practical advantages that it offers. As well as representing a simple non-technical solution, it means that each developmental analysis does not have to be replicated for both teacher and parent data. However, when parents and teachers report such different trajectories as observed in the current study, discarding information from one informant will give an incomplete picture of developmental trends. There is also no necessary reason to believe that either teacher or parent ratings are more suitable as informants; making the choice of which informant to rely on difficult. In fact, important advantages and disadvantages have been identified for both. (e.g. Eaves et al., 1997; Hartman et al., 2007; Jackson & King, 2004; Loeber et al., 1990).

It is also worthwhile comparing the correlated factors approach used in this study to other proposed latent variable approaches in which item variance is partitioned into that which is common to informants and that which is unique to one informant. Bauer et al.

(2013), for example, proposed a trifactor model for multi-informant data. In this model, each item loads on three factors: a common perspective factor, an informant perspective factor and an item specific factor (see Martel et al., 2017 for an application in ADHD). The main disadvantage of this approach is that measurement models of this complexity can show convergence difficulties (e.g. Maydeu-Olivares & Coffman, 2006) and are likely to be difficult to combine with developmental trajectory analyses, especially growth mixture models. Two-step approaches whereby factor scores are estimated in a first-step from a tri-factor model and then analysed in the developmental model in a second step are possible in principle. However, whether factor scores with sufficiently high determinacies could be obtained for all factors is questionable, especially for the common perspective factor in cases of low inter-informant agreement. Thus, the trifactor model may be useful for gaining insights into cross-informant agreement and as a measurement model for simple analyses. Its utility in more complex analyses may be more limited.

Finally, it is important to note the limitations of the current study. First, our measure of ADHD symptoms was - though psychometrically supported - brief. Second, we had only three time points of common parent- and teacher data and these were restricted to the elementary school years. Future studies may be able to establish whether cross-informant agreement on development is similar across longer time spans or different developmental periods. Past adolescence, for example, self- and peer- ratings are often used and it would be valuable to investigate the extent to which these concur with each other and with parent and teacher ratings. Beyond elementary school, teacher ratings may become less reliable because at this stage teachers spend less concentrated periods of time with any given student. On the other hand, peer- and self- ratings may become relatively more reliable. Third, we used a normative sample. Both clinical and normative samples are required to provide the full picture of phenotypes such as ADHD symptoms that show meaningful variation both above

and below clinical thresholds. Normative samples are useful for obtaining evidence on ADHD symptom trajectories in the population and avoiding potential ascertainment biases and related problems associated with using clinical samples (e.g. Berkson's bias, range restriction). However, replication in clinical samples will be important to establish whether clinically diagnosed individuals show similar levels and pattern of informant (dis-)agreement in symptom trajectories over time.

Conclusions

Cross-informant agreements were in the low to moderate range for ratings of ADHD symptom developmental trajectories. In the case of inattention, opposite developmental trends were suggested by parent- versus teacher- reports. Our results underline the importance of incorporating multiple raters in developmental analyses of ADHD symptoms and suggest that parent- and teacher- ratings should be analysed separately. Specifically, for developmental analyses, we recommend analysing multi-informant data using separate factors in, for example, a parallel process model. This approach allows different patterns of results to emerge for different informants while also acknowledging the correlation in the data they provide. Further, this approach allows for numerous extensions that allow for further investigations of the source and implications of differences between informants.

References

Achenbach, T. M. (2006). As others see us: Clinical and research implications of cross-informant correlations for psychopathology. *Current Directions in Psychological Science*, *15*, 94-98.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Antrop, I., Roeyers, H., Oosterlaan, J., & Van Oost, P. (2002). Agreement between parent and teacher ratings of disruptive behavior disorders in children with clinically diagnosed ADHD. *Journal of Psychopathology and Behavioral Assessment*, *24*, 67-73.

Arnold, L. E., Ganocy, S. J., Mount, K., Youngstrom, E. A., Frazier, T., Fristad, M., ... & Demeter, C. (2014). Three-year latent class trajectories of attention-deficit/hyperactivity disorder (ADHD) symptoms in a clinical sample not selected for ADHD. *Journal of the American Academy of Child & Adolescent Psychiatry*, *53*, 745-760.

Bauer, D. J., Howard, A. L., Baldasaro, R. E., Curran, P. J., Hussong, A. M., Chassin, L., & Zucker, R. A. (2013). A trifactor model for integrating ratings across multiple informants. *Psychological Methods*, *18*, 475.

De Los Reyes, A. (2011). Introduction to the special section: More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, *40*, 1-9.

De Los Reyes, A. (2013). Strategic objectives for improving understanding of informant discrepancies in developmental psychopathology research. *Development and Psychopathology*, *25*, 669-682.

De Los Reyes, A., Henry, D. B., Tolan, P. H., & Wakschlag, L. S. (2009). Linking informant discrepancies to observed variations in young children's disruptive behavior. *Journal of Abnormal Child Psychology*, *37*, 637-652.

Döpfner, M., Hautmann, C., Görtz-Dorten, A., Klasen, F., Ravens-Sieberer, U., & BELLA Study Group. (2015). Long-term course of ADHD symptoms from childhood to early adulthood in a community sample. *European Child & Adolescent Psychiatry*, *24*, 665-673.

Eaves, L. J., Silberg, J. L., Meyer, J. M., Maes, H. H., Simonoff, E., Pickles, A., ... & Neale, M. C. (1997). Genetics and developmental psychopathology: 2. The main effects of genes and environment on behavioral problems in the Virginia Twin Study of Adolescent Behavioral Development. *Journal of Child Psychology and Psychiatry*, *38*, 965-980.

Eisner, M., & Ribeaud, D. (2007). Conducting a criminological survey in a culturally diverse context: Lessons from the Zurich Project on the Social Development of Children. *European Journal of Criminology*, *4*, 271-298.

Eisner, N., Murray, A.L., Eisner, M., Ribeaud. (2018). An Analysis of Non-response and Attrition in the Zurich Project on Social Development from Childhood to Adulthood (z-proso). *International Journal of Behavioral Development.* In press.

Faraone, S. V., Biederman, J., & Mick, E. (2006). The age-dependent decline of attention deficit hyperactivity disorder: A meta-analysis of follow-up studies. *Psychological Medicine, 36,* 159-165.

Gershon, J., & Gershon, J. (2002). A meta-analytic review of gender differences in ADHD. *Journal of Attention Disorders*, *5*, 143-154.

Groen-Blokhuis, M. M., Middeldorp, C. M., Kan, K. J., Abdellaoui, A., van Beijsterveldt, C. E., Ehli, E. A., ... & Hottenga, J. J. (2014). Attention-deficit/hyperactivity disorder polygenic risk scores predict attention problems in a population-based sample of children. *Journal of the American Academy of Child & Adolescent Psychiatry*, *53*, 1123-1129.

Hart, E. L., Lahey, B. B., Loeber, R., Applegate, B., & Frick, P. J. (1995). Developmental change in attention-deficit hyperactivity disorder in boys: a four-year longitudinal study. *Journal of Abnormal Child Psychology*, *23*, 729-749.

Hartman, C. A., Rhee, S. H., Willcutt, E. G., & Pennington, B. F. (2007). Modeling rater disagreement for ADHD: are parents or teachers biased?. *Journal of Abnormal Child Psychology*, *35*, 536-542.

Imeraj, L., Antrop, I., Roeyers, H., Deboutte, D., Deschepper, E., Bal, S., & Sonuga-Barke, E. (2016). The impact of idle time in the classroom: Differential effects on children with ADHD. *Journal of Attention Disorders*, *20*, 71-81.

Jackson, D. A., & King, A. R. (2004). Gender differences in the effects of oppositional behavior on teacher ratings of ADHD symptoms. *Journal of Abnormal Child Psychology*, *32*, 215-224.

Jester, J. M., Nigg, J. T., Adams, K., Fitzgerald, H. E., Puttler, L. I., Wong, M. M., & Zucker, R. A. (2005). Inattention/hyperactivity and aggression from early childhood to adolescence: Heterogeneity of trajectories and differential influence of family environment characteristics. *Development and Psychopathology*, *17*, 99-125.

Kennerley, S., Jaquiery, B., Hatch, B., Healey, M., Wheeler, B. J., & Healey, D. (2016). Informant Discrepancies in the Assessment of Attention-Deficit/Hyperactivity Disorder. *Journal of Psychoeducational Assessment*, 0734282916670797.

Lahey, B. B., Pelham, W. E., Loney, J., Lee, S. S., & Willcutt, E. (2005). Instability of the DSM-IV subtypes of ADHD from preschool through elementary school. *Archives of General Psychiatry, 62*, 896-902.

Laird, R. D., & De Los Reyes, A. (2013). Testing informant discrepancies as predictors of early adolescent psychopathology: Why difference scores cannot tell you what you want to know and how polynomial regression may. *Journal of Abnormal Child Psychology, 41*, 1-14.

Litson, K., Geiser, C., Burns, G. L., & Servera, M. (2016). Trait and state variance in multi-informant assessments of ADHD and academic impairment in Spanish first-grade children. *Journal of Clinical Child & Adolescent Psychology*, 1-14.

Loeber, R., Green, S. M., & Lahey, B. B. (1990). Mental health professionals' perception of the utility of children, mothers, and teachers as informants on childhood psychopathology. *Journal of Clinical Child Psychology*, *19*, 136-143.

Malone, P. S., Van Eck, K., Flory, K., & Lamis, D. A. (2010). A mixture-model approach to linking ADHD to adolescent onset of illicit drug use. *Developmental Psychology*, *46*, 1543.

Martel, M. M., Nigg, J. T., & Schimmack, U. (2017). Psychometrically Informed Approach to Integration of Multiple Informant Ratings in Adult ADHD in a Community-Recruited Sample. *Assessment*, *24*, 279-289

Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological Methods*, *11*, 344.

Moffitt, T. E., Houts, R., Asherson, P., Belsky, D. W., Corcoran, D. L., Hammerle, M., ... & Poulton, R. (2015). Is adult ADHD a childhood-onset neurodevelopmental disorder? Evidence from a four-decade longitudinal cohort study. *American Journal of Psychiatry*, *172*, 967-977.

Murray, A.L**.,** Booth, T., Eisner, M., Auyeung., B., Murray, G., Ribeaud, D. (2018). Sex differences in ADHD trajectories across childhood and adolescence. Manuscript submitted for publication.

Murray, A. L., Eisner, M., Obsuth, I., & Ribeaud, D. (2017). Identifying Early Markers of "Late Onset" Attention Deficit and Hyperactivity/Impulsivity Symptoms. *Journal of Attention Disorders*, 1087054717705202.

Murray, D. W., Kollins, S. H., Hardy, K. K., Abikoff, H. B., Swanson, J. M., Cunningham, C., ... & McCracken, J. T. (2007). Parent versus teacher ratings of attention-deficit/hyperactivity disorder symptoms in the Preschoolers with Attention-Deficit/Hyperactivity Disorder Treatment Study (PATS). *Journal of Child and Adolescent Psychopharmacology*, *17*, 605-619.

Murray, A. L., Obsuth, I., Eisner, M., & Ribeaud, D. (2017). Evaluating Longitudinal Invariance in Dimensions of Mental Health Across Adolescence: An Analysis of the Social Behavior Questionnaire. *Assessment*, 1073191117721741.

Murray, A.L., Eisner, M., & Ribeaud, D. (2017). Can the Social Behavior Questionnaire help meet the need for dimensional, transdiagnostic measures of childhood and adolescent psychopathology?. *European Journal of Psychological Assessment*. In press.

Murray, A. L., Ribeaud, D., Eisner, M. (2018). Should we subtype ADHD according to the context in which symptoms occur? Criterion validity of context-based ADHD subtypes. Manuscript submitted for publication.

Muthén, L. K., & Muthén, B. O. (2014). *Mplus User's Guide*. (7th ed.). Muthén & Muthén, Los Angeles.

Narad, M. E., Garner, A. A., Peugh, J. L., Tamm, L., Antonini, T. N., Kingery, K. M., ... & Epstein, J. N. (2015). Parent–teacher agreement on ADHD symptoms across development. *Psychological Assessment*, *27*, 239.

Pas, E. T., & Bradshaw, C. P. (2014). What affects teacher ratings of student behaviors? The potential influence of teachers' perceptions of the school environment and experiences. *Prevention Science, 15,* 940-950.

Pingault, J. B., Tremblay, R. E., Vitaro, F., Carbonneau, R., Genolini, C., Falissard, B., & Côté, S. M. (2011). Childhood trajectories of inattention and hyperactivity and prediction of educational attainment in early adulthood: a 16-year longitudinal population-based study. *American Journal of Psychiatry*, *168*, 1164-1170.

Polanczyk, G., de Lima, M. S., Horta, B. L., Biederman, J., & Rohde, L. A. (2007). The worldwide prevalence of ADHD: a systematic review and metaregression analysis. *American Journal of Psychiatry*, *164*(6), 942-948.

Sollie, H., Larsson, B., & Mørch, W. T. (2013). Comparison of mother, father, and teacher reports of ADHD core symptoms in a sample of child psychiatric outpatients. *Journal of Attention Disorders*, *17*, 699-710.

Sayal, K., & Goodman, R. (2009). Do parental reports of child hyperkinetic disorder symptoms at school predict teacher ratings?. *European Child & Adolescent Psychiatry, 18,* 336-344.

Sonuga-Barke, E. J., Van Lier, P., Swanson, J. M., Coghill, D., Wigal, S., Vandenberghe, M., & Hatch, S. (2008). Heterogeneity in the pharmacodynamics of two long-acting methylphenidate formulations for children with attention deficit/hyperactivity disorder. *European Child & Adolescent Psychiatry*, *17*, 245-254.

Swanson, J. M., Hinshaw, S. P., Arnold, L. E., Gibbons, R. D., Marcus, S. U. E., Hur, K., ... & Hechtman, L. (2007). Secondary evaluations of MTA 36-month outcomes: propensity score and growth mixture model analyses. *Journal of the American Academy of Child & Adolescent Psychiatry*, *46*, 1003-1014.

Tremblay, R. E., Loeber, R., Gagnon, C., Charlebois, P., Larivee, S., & LeBlanc, M. (1991). Disruptive boys with stable and unstable high fighting behavior patterns during junior elementary school. *Journal of Abnormal Child Psychology*, *19*, 285-300.

Van Lier, P. A., Der Ende, J. V., Koot, H. M., & Verhulst, F. C. (2007). Which better predicts conduct problems? The relationship of trajectories of conduct problems with ODD and ADHD symptoms from childhood into adolescence. *Journal of Child Psychology and Psychiatry*, *48*, 601-608.

Van Der Ende, J., & Verhulst, F. C. (2005). Informant, gender and age differences in ratings of adolescent problem behaviour. *European Child & Adolescent Psychiatry, 14*, 117-126.

Willcutt, E. G., Nigg, J. T., Pennington, B. F., Solanto, M. V., Rohde, L. A., Tannock, R., ... & Lahey, B. B. (2012). Validity of DSM-IV attention deficit/hyperactivity disorder symptom dimensions and subtypes.

Yeguez, C. E., & Sibley, M. H. (2016). Predictors of Informant Discrepancies Between Mother and Middle School Teacher ADHD Ratings. *School mental health*, *8*(4), 452-460.

Youngstrom, E., Loeber, R., & Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *Journal of consulting and clinical psychology*, *68*(6), 1038.

Tables

Table 1:

ADHD items

Item number	Domain	Item content
10	Hyperactivity/impulsivity	<CHILD> is impulsive, acts without thinking.
11	Hyperactivity/impulsivity	<CHILD> has difficulty awaiting turn in games or groups.
12	Hyperactivity/impulsivity	<CHILD> can't sit still, is restless, or hyperactive.
13	Hyperactivity/impulsivity	<CHILD> fidgets.
14	Inattention	<CHILD> cannot settle to anything for more than a few moments.
15	Inattention	<CHILD> is distractible, has trouble sticking to any activity.
16	Inattention	<CHILD> can't concentrate, can't pay attention for long.
17	Inattention	<CHILD> is inattentive.

Table 2:

Unstandardised parameter estimates for inattention parallel process model

	Intercept factor mean (SE)	**Slope factor mean (SE)**	**Intercept-slope covariance (SE)**
Parent	.00 (N/A)	.44 (0.13)	-0.24 (0.35)
Teacher	.56 (0.15)	-.44 (0.14)	-1.93 (0.85)

Table 3:

Unstandardised parameter estimates for hyperactivity/impulsivity parallel process model

	Intercept factor mean (SE)	**Slope factor mean (SE)**	**Intercept-slope covariance (SE)**
Parent	.00 (N/A)	-0.08 (0.22)	0.47 (0.57)
Teacher	-1.55 (0.56)	0.21 (0.46)	-5.98 (3.90)

Figure Captions

Figure 1:

Parent- and teacher-reported average trajectories

Conflict of Interest

The authors have no conflicts of interest to declare.